Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Selection on DRD4 haplotypes in a natural great tit population in relation to personality

Bimal Chakkingal Bhaskaran

Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics

458

Uppsala 2009

Master's Thesis, 30 hp

Erasmus Mundus Programme – European Master in Animal Breeding and Genetics

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Selection on DRD4 haplotypes in a natural great tit population in relation to personality

Bimal Chakkingal Bhaskaran

Supervisors:

Richard Crooijmans, WUR, Animal Breeding and Genetics Group
Kees van Oers, NIOO, Animal Ecology
Susanne Kerje, SLU, Department of Animal Breeding and Genetics

Examiner:

Birgitta Malmfors, SLU, Department of Animal Breeding and Genetics

Credits: 30 HEC
Course title: Degree project in Animal Science
Course code: EX0556
Programme: Erasmus Mundus programme
– European Master in Animal Breeding and Genetics
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2009
Name of series: Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics, 458
On-line publication: http://epsilon.slu.se

Key words: DRD4, dopamine receptor, Great tit, early exploratory behaviour

Education and Culture
Erasmus Mundus

SELECTION ON DRD4 HAPLOTYPES IN A NATURAL GREAT TIT POPULATION IN RELATION TO PERSONALITY

BIMAL CHAKKINGAL BHASKARAN
791106068090

THESIS ANIMAL BREEDING AND GENETICS (ABG 804030)
JUNE 2009

WAGENINGEN UR
For quality of life

Animal Breeding and Genetics

SUPERVISORS
Richard Crooijmans (ABGC)
Kees van Oers (NIOO)
Susanne Kerje (SLU)

Education and Culture
Erasmus Mundus

Acknowledgement

I take this opportunity to extend my deep sense of gratitude and words of appreciation towards those who helped me during the pursuit of my present study.

I express my deepest gratitude and sincere thanks to my supervisor Dr. Richard Crooijmans for his invaluable guidance, persistent encouragement and active persuasion through out the course of my study. Words of special appreciation and thankfulness would fail to express my deepest sense of gratitude to Dr. Kees van Oers. I am highly obliged and grateful to Dr. Hendrik-Jan Megens, Prof. Dr. Martien Groenen and Dr. Susanne Kerje for their valuable suggestions and unstinted cooperation extended during the entire period of my work. I wish to put on records my sincere thanks and esteemed gratitude to Bert Dibbits, Sylvia Kinders and Tenaeka for helping me with the lab work and making themselves available for clearing my doubts.

This work would have not seen the light of the day, had I not been constantly inspired and helped by the people around me. I wish to express my sincere gratitude and thanks to my friends Dinesh, Preethi, Naveen, Albin, Geo, Bijo, Sushi, Biju, Yalda, Chris and Roopika for their collective spirit, ever-willing help and goodwill.

Finally, no words of admiration can reply the love, inspiration, encouragement and constant support of my father, mother, sister, brother-in law and other family members at the times of pain, frustration, hurdles and happiness.

Summary

Variations in neurotransmitter-related genes are reported to be associated with personality traits among humans. One of the genes, the dopamine receptor (*Drd4*) gene showed a relation with novelty-seeking behaviour or curiosity traits. Moreover, in human, the dopamine receptor is the target site for drugs used in treating Parkinson's disease and schizophrenia. Non human vertebrates and free living species can provide better understanding of the genotype personality relationships as they can be measured under standardized selection experiments. The great tit (*Parus major*) is one such model species used in these types of studies. Temperamental traits are heritable as well as linked to fitness traits, which makes it important in the study of ecology and evolution. A recent study by Fidler et al (2007) detected 73 polymorphisms (66 SNPs and 7 indels) in the great tit Drd4 orthologue (GenBank: DQ006801.1) The objectives of the current study were i) to amplify and sequence selected regions in dopamine receptor gene in two lines (slow and fast) of great tit, including from the wild and ii) to identify SNPs and haplotypes within this gene. iii) To develop a strategy for typing the different haplotypes within a large population (> 1000 animals).

Two different lines of a great tit population, selected for slow and fast Early Exploratory Behaviour (EEB) were considered for the experiment. These birds were reared under captive conditions at Netherlands Institute of Ecology (NIOO). A total of 19 birds from the fast line and 21 birds from the slow line were used in the study. Apart from the captive population of these two lines, a wild population (N=10) representing an out-group was also tested to identify the pattern followed under natural selection. Twelve regions within the dopamine receptor gene (Fig 1) were selected based on either SNP density or their proximity to indels

Six haplotype blocks were identified within the gene. The SNPs constituting these blocks had, on an average, a MAF of 0.2 and were in high LD, which would eventually make it easier to find them and thereby enable to genotype a larger population using these six haplotype blocks. A significant association of SNPs 79 and 81 with the slow phenotype was observed, suggestive of a region which could be in association with this trait. SNP 76, which was reported (Fidler et al, 2007) to be associated with novelty seeking behaviour was found to be not significant. However, these SNPs are in close correlation (r2=0.69) with SNP 76 and hence indicate a region of strong association with the trait. The effect of introns, rather than the coding regions, in gene regulation could be the possible reason for this strong association. Further, a low level of LD within the gene supports the speculation that the causative mutation is within the dopamine receptor gene. The results from the wild out group weren't significant owing to their small number but a highly similar trend was noticed suggestive of an association with the trait. A more detailed study could explain the trends followed in natural selection and evolution.

Table of contents

1 Introduction

Variation in neurotransmitter-related genes has been reported to be associated with personality traits among humans. Dopamine being one of the most important neurotransmitters in the brain, the gene coding it- the Dopamine receptor (*Drd4*) has been the subject of numerous candidate-gene studies in psychiatry. Also, due to various factors like its specific location in limbic areas of brain, several polymorphisms in the coding regions and their effect on ligand binding properties of its protein, it is one of the most studied genes in behavioural science (Reviewed in Galenter *et al*, 1997). Moreover, dopamine receptor is the target site for drugs used in treating Parkinson's disease and schizophrenia (Rang *et al*, 2001). Several studies indicated *Drd4* gene to be associated with novelty-seeking behaviour or curiosity traits (Reviewed in Kluger *et al*, 2002). But, there are conflicting results about the role of *Drd4* in novelty seeking in human. These are considered mainly due to either a weak effect, an association found only in certain populations, or a false positive resulting from population stratification (Reviewed in Paterson *et al*, 1998). It is in this scenario the importance of non human vertebrates and free living species come in to the picture.

Genetic studies on personality differences in wild species are relatively rare. But, studying 'within population' animal personality differences is important to understand ecology and evolution. This is because of the fact that temperamental traits are heritable and are linked to fitness and other traits important in ecology and evolution. (Réale *et al*, 2007). Moreover, variation in personality traits is assumed to be the product of natural selection (Drent *et al*, 2003). Although genetic studies on human personalities are very valuable, it is a big hurdle to explain the behavioural variations in an evolutionary point of view (Reviewed in van Oers *et al*, 2005). Non human vertebrates and free living species can provide better understanding of the genotype personality relationships as it can be measured under standardized selection experiments. The influence of culture and environment has less significant role in these groups, when compared to human (Fidler *et al*, 2007). There exists highly consistent individual variation in personalities, which allows us to measure the behaviour under standardized conditions on birds bred in

captivity. These standardized measurements can then be further linked to the behaviour under natural conditions and thereby measure natural selection in the field. (Drent *et al*, 2003).

The great tit (*Parus major*) is a classical model species used in behaviour studies. Their well known behaviour ecology, similarity to other species in behavioral patterns, established selection lines with respect to behaviour and also the easiness to rear them in captivity makes it a primary choice in ecological research (Reviewed in Groothuis and Carere, 2005). Considerable amounts of genetic variations for personality traits are found in great tits. Variations in early exploratory behaviour (EEB) within selected lines of great tit are attributed to their wild caught parents (Drent *et al*, 2003). Hence, finding the genetic basis for these variations in captive and wild populations can explain trends in natural selection to a large extent.

Confirmation for the involvement of any candidate gene in expression of a phenotypic trait requires identification of polymorphisms in that gene which is significant statistically. Due to the advances in molecular genetics, it is now possible to sequence genes of interest to find variations even at single nucleotide level. Single Nucleotide Polymorphisms (SNPs) has been increasingly used as genetic markers in molecular studies of ecology and evolution. It has several advantages over other markers due to various reasons (Reviewed in Berlin *et al*, 2008). Moreover, they can be used in constructing haplotypes and also in Linkage Disequilibrium (LD) studies. Use of more conservative statistical criteria for significance, employing gene haplotypes, as well as LD studies might be useful to rectify the inconclusive results associated with genotype personality association studies.

A recent study by Fidler *et al* (2007) detected 73 polymorphisms (66 SNPs and 7 indels) in the great tit Drd4 orthologue (GenBank: DQ006801). They found significant association of the polymorphism (Drd4 SNP830) with EEB in two selection lines (a slow and a fast) and a wild population of great tit. But, no further information is available on other SNPs and indels and their association with these phenotypes. Hence the present

study aimed at validating all the SNPs together with finding all the SNPs present within the *Drd4* gene and to construct haplotypes in wild and captive populations of great tit. An association study was also done using all the information from SNPs, indels and haplotypes. Testing a wild population together with the captive populations would possibly help us to understand the trend followed under natural selection and explain evolution.

The objectives of the current study were as follows

1. To amplify and sequence selected regions in dopamine receptor gene in two lines (slow and fast) of great tit, including from the wild to validate the SNPs.

2. Identify additional SNPs and construct haplotypes within this gene.

3. Develop a strategy for typing the different haplotypes within a large population (> 1000 animals) of great tit.

2 Materials and Methods

2.1 Birds used in the study

Two different lines of a great tit population, selected for slow and fast Early Exploratory Behaviour (EEB) were considered for the experiment. The birds were reared under captive conditions at Netherlands Institute of Ecology (NIOO). These lines started from birds caught from the wild in early nineties and were selected for 4 generations and further maintained. A total of 19 unrelated birds from the fast line and 21 unrelated birds from the slow line were used in the study. Apart from the captive population of these two lines, a wild population (N=10) representing an out-group was also tested to identify the pattern followed under natural selection.

2.2 Primer design, PCR and Sequencing

Twelve regions within the dopamine receptor gene (10897bp, GenBank acc. no. DQ006801) were selected based on the SNP density and their proximity to indels, thereby enabling to type the indels together with the SNPs encompassing them (Fig 2.1)

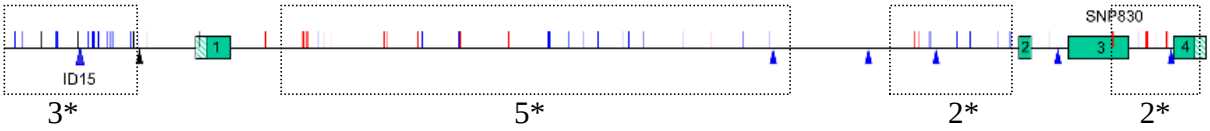


Fig 2.1: Schematic representation of the *P. major Drd4* gene structure. Exons are shown as green boxes (coding regions full colour, untranslated regions striped), SNPs as vertical lines and indels as triangles. Indel 1(ID 15) and SNP 830 are also marked (Fidler *et al*, 2007)

* = No of regions selected within each block

Primers for PCR were designed by Primer3 (v. 0.4.0) (Steve Rozen and Helen J. Skaletsky, 2000) using the *Drd4* gene sequence as template and then amplification of these selected regions were done. The regions selected included three from the 5'UTR, seven from intron 1 and two regions encompassing third and fourth exon. Primer details and the description of region amplified are given in Appendix A.

DNA samples from the birds selected for the study were obtained from NIOO. Initially, all the primers were tested on two samples and the whole procedures up to the sequencing steps were standardized. All regions (except Product one) were amplified by

routine PCR. PCR amplifications were performed in 12µl reactions in an Applied Biosystems Gene Amp PCR System 9700 thermal cycler. Individual mixes contained approximately 60 ng DNA template, 2x PCR Master Mix containing 1.5mM $MgCl_2$ (Thermo Scientific ABgene ® UK) and 2.4µM of each primer. PCR profiles consisted of 5 min denaturation at 95 ^{o}C followed by 35 cycles of 30 sec denaturation at 95 ^{o}C, 45 sec annealing at 55-60 ^{o}C and 90 sec extension at 72 ^{o}C with a final 10 min 72 ^{o}C step (See Appendix B for PCR conditions). Hot-start PCR for product one was done using a Hot-start Master Mix (Hst). A single reaction of 20 µl contained 10 µl Hst, Q solution - 4 µl, 0.4 µl MQ, 1.6 µl primer mix* and 4 µl genomic DNA (10ng/ µl). {* Primer mix = 4 µl Forward primer (40 µM/ µl) + 4 µl Reverse primer (40 µM/ µl) + 16 µl MQ, mixed well and taken 1.6 µl }.

All products except indel 3 were sequenced by BigDye Terminator 3.1 (Sanger sequencing) using ABI 3730 DNA Analyzer (Applied Biosystems). A 48bp long indel (indel 3) was typed after amplifying the region and running on a 2% agar gel (Appendix C). Sequences were analyzed using Staden package (Staden et al. 2000) and the segments were scanned for Single Nucleotide Polymorphisms. Indels were also typed based on its presence (+/+) or absence in either one (+/-) or both strands (-/-) of DNA.

2.3 Haplotype and Linkage Disequilibrium (LD) Plot Construction

SNP information from all individuals for all the loci was retrieved and then an input file for the programme Phase (v 2.1) (Stephens et al. 2001 and Stephens and Scheet, 2005) was created.

The default structure for the phase input file is represented as follows:

Number of Individuals

Number of Loci

P Position (1) Position (2) Position (Number of Loci)

Locus Type (1) Locus Type (2) ... Locus Type (Number of Loci)

ID (1)

Genotype (1)

ID (2)

Genotype (2).

where the quantities above are as follows:

1. Number of individuals - An integer specifying the number of individuals who have been genotyped.

2. Number of Loci - An integer specifying the number of loci or sites at which each individual has been typed.

3. P - The character 'P' (upper case, without quotation marks).

4. Position (i) - A number indicating the position of locus i, relative to some arbitrary reference point. The loci must be in their physical order along the chromosome (i.e. these Positions must be increasing).

5. Locus Type (i) - A letter indicating the type of locus i. The options are (a) S for a biallelic (SNP) locus, or biallelic site in sequence data. (b) M for microsatellite, or other multi-allelic locus (e.g. tri-allelic SNP, or HLA allele). These characters can be separated by spaces, if desired. In this study, only SNP information was there. All the indel information was converted to SNP format.

6. ID (i) - A string, giving a label for individual I, ID (Number of Individuals)

Genotype (Number of Individuals)

7. Genotype (i) - The genotypes for the i'th individual. This is given on two consecutive rows. At each locus, one allele is entered on the first row, and one on the second row. It does not matter which allele is entered on each row. For biallelic loci, any two characters (e.g. A/C, G/T, 0/1) can be used to represent the two alleles, and they do not need to be separated by a space. Missing alleles at SNP loci should be entered as '?'

Genotype information was eventually used for creating all possible haplotypes among the individuals. The output from Phase was then used for constructing cluster dendrogram using R script, (kindly supplied by Hendrik-Jan Megen). Further, an input file for Haploview (version 4.1) (Barrett et al, 2005) was made in the linkage format. For linkage format, data was given in the Linkage Pedigree format, with columns of family, individual, father, mother, gender, affected status and genotypes. The default structure of the Haploview Linkage format is as follows

(a) Pedigree name - A unique alphanumeric identifier for this individual's family. Unrelated individuals should not share a pedigree name.

(b) Individual ID An alphanumeric identifier for this individual. Should be unique within his family

(c) Father's ID- Identifier corresponding to father's individual ID or "0" if unknown father.

(d) Mother's- ID Identifier corresponding to mother's individual ID or "0" if unknown mother

(e) Sex - Individual's gender (1=MALE, 2=FEMALE).

(f) Affection status - Affection status to be used for association tests (0=UNKNOWN, 1=UNAFFECTED, 2=AFFECTED).

(g) Marker genotypes - Each marker is represented by two columns (one for each allele, separated by a space) and coded either ACGT or 1-4 where: 1=A, 2=C, 3=G, T=4. A 0 in any of the marker genotype position indicates missing data.

Haplotype blocks and LD plot was created for all the three groups combined and also within individual groups. Mean squared correlation in allelic state between pairs of SNPs (r^2) and D' (the difference between the observed and the expected gametic frequencies standardized by the theoretical maximum for the observed allele frequencies) were evaluated to estimate the level of LD.

2.4 Association studies

An association study was performed using all the SNPs, indels and haplotype blocks found in the study, using Haploview programme. Analysis was done for all the phenotypic groups jointly and separately. Statistical significance were confirmed at $P<0.01$ and $P<0.05$, after 10000 permutations, during the analysis.

3 Results

A total of 80 SNPs and 5 indels were found upon sequencing these selected regions. Out of this, 34 SNPs and one indel were newly found and not previously reported in the study by Fidler and co-workers (2007). Complete information about SNPs and indels newly found as well as previously known and validated by this study are given in Appendix D. Genotype information from these regions was utilized to construct haplotypes using the programme Phase. There were 82 unique haplotypes found using Phase (Appendix E). The output from Phase was then utilized to construct cluster dendrogram (Appendix F).

From the result, it was evident that these haplotypes were distributed in both the groups used in the study. The wild out-group also showed a similar trend although no definite haplotype block was observed. Linkage Disequilibrium (LD) plot and haplotype blocks were constructed using genotype information from all the 3 groups, by means of Haploview programme and it showed 6 haplotype blocks (Fig 3.1 and 3.2).

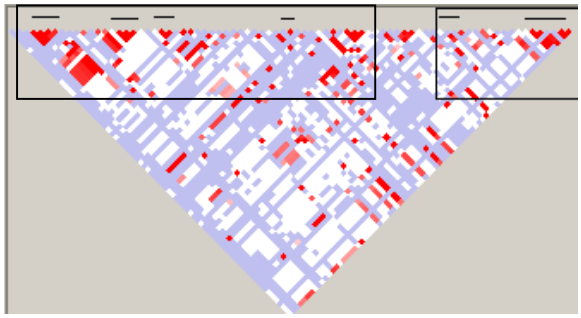


Fig 3.1: LD plot showing 6 haplotype blocks from all the 3 groups combined

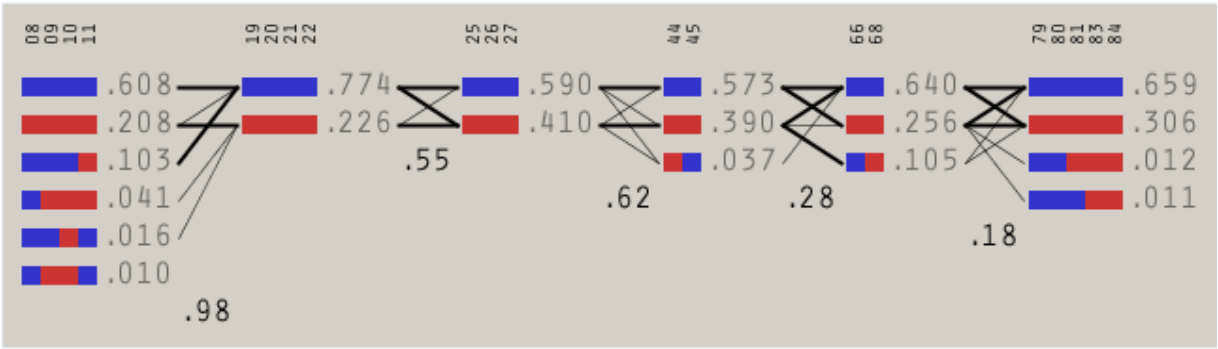


Fig 3.2: Showing 6 haplotype blocks using genotype information from all three groups

RESULTS

On average, the minor allele frequencies (MAF) of the SNPs constituting these blocks were above 0.2 (Appendix G). Further, LD plot and haplotype blocks were constructed separately for these groups which showed 3 blocks for the fast line, 5 for the slow line and none for the wild out group. (Fig 3.3 and 3.4).The SNPs in these blocks had a high correlation coefficient (r^2) and D'value. But there was very little LD evident within this gene. Block 1 and Block 4 were unique for the slow line and part of block 2 was unique for the fast line. Analysis using Haploview revealed significant association ($p<0.01$, 10000 permutations) of phenotype (slow EEB) with SNPs 79 and 81 located in the third introns (Appendix H). The sixth haplotype block which included these SNPs were also significantly associated with this phenotype. A similar trend was seen in the wild group but no significant association was observed.

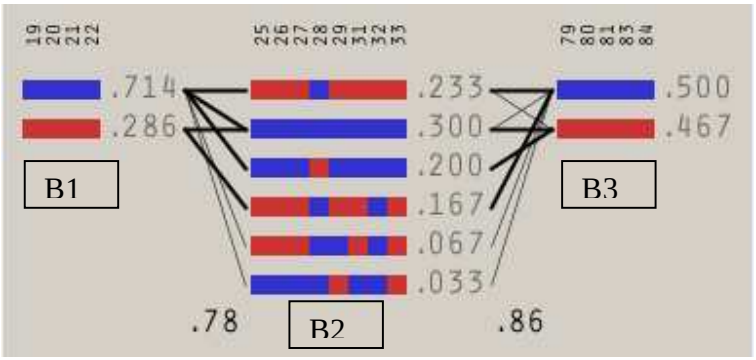


Fig 3.3: Showing 3 haplotype blocks for the fast line; B1 to B3- Haplotype blocks

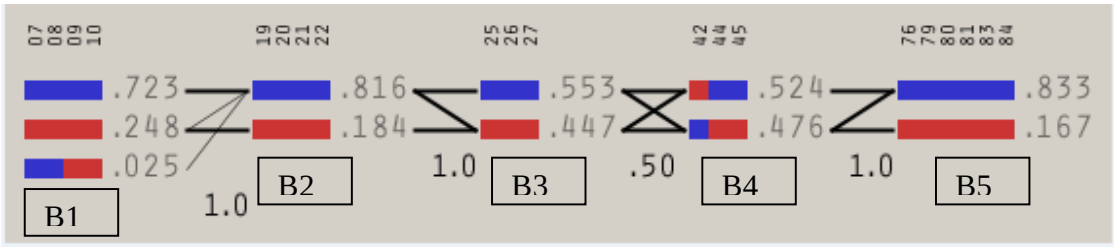


Fig 3.4: Showing 5 haplotype blocks for the slow line; B1 to B5- Haplotype blocks

4 Discussion

A total of 80 SNPs and 5 indels were found in this study. Majority of the SNPs in these regions (Fidler *et al*, 2007) were validated upon sequencing. Apart from that, a total of 34 SNPs and one indel were newly found and not previously reported by Fidler and co-workers (2007) (Appendix D). Finding more SNPs from this region certainly helped in improving the haplotype resolution. The SNP information was later used for haplotype construction.

Using Phase, 82 unique haplotypes (Appendix E) were constructed, which suggested a scattered distribution of variation in the groups studied. But, a cluster dendrogram (Appendix F) made using the output from Phase gave insights about the closeness of these haplotypes. There were small clusters of haplotypes, separated by a few bases. This prompted for the construction of haplotype blocks and LD plot using Haploview.

Six haplotype blocks were identified within the gene. The SNPs constituting these blocks had, on an average, a minor allele frequency (MAF) above 0.2 (Appendix G) and were in high LD. A relatively high MAF would make it possible to find them with ease and thereby enable to genotype a larger population using these six haplotype blocks. A representative SNP from each block could be selected and typed for genotyping a larger population.

Although unique haplotypes were seen in both fast and slow lines (Fig 3.3 and 3.4), no significant association of these blocks with the respective lines was found. However, a significant association of SNPs 79 and 81 with the slow phenotype was observed, suggestive of a region which could be in association with this trait. SNP 76 (Drd4 SNP 830), which was reported (Fidler et al, 2007) to be associated with novelty seeking behaviour was found to be not significant in this study. However, these SNPs were in close correlation (r^2=0.69) with SNP 76 and hence indicate a region of strong association with the trait. Also, the haplotype block (Block 5 in slow line, Fig 3.4) representing these SNPs were significantly (P<0.01) associated with the slow phenotype. Further, a low

level of LD within the gene supports the speculation that the causative mutation is within the dopamine receptor gene.

A notable finding from these results is the significant association of SNPs located in the intron. In their study, Fidler and coworkers (2007) followed the SNP in the third exon and also one indel in the promoter region. The results from this study strongly suggest that it is not the SNP in exon 3 but some regulatory region in the intron 3 which is involved in the phenotypic variation. This was in accordance with the study by Fidler *et al*. (2007), where they found a synonymous mutation (SNP 830) indicating that the actual functional mutation is some where else and linked to this SNP. An effect of intron in the regulatory mechanism was probably not measured because introns were considered junk DNA at that point of time. But there exists many evidences which proves that the introns have variety of functions, including for regulation and structural purposes, and that many of the roles now hypothesized for introns are plausible but need further elucidation (Wang and Christopher, 2008., Brudno *et al*, 2001., Dietrich *et al*, 2001) . Hence, it would be really interesting to look deeply in to these regions for finding any possible alternate splicing or regulatory sites.

It is very interesting to find that a lot of variations are still maintained in these selection lines even after many generations of captivity and selection. Similar trend seen in the wild population suggests that it is the same force of genetic selection being followed in the nature. It gives a strong evidence for natural selection for this behavioural trait. Even though the results from the wild out group weren't significant owing to their small number, a highly similar trend was noticed suggestive of an association with the trait. A significant association of these polymorphisms couldn't be seen initially when the phenotypic data from the wild out-group was not used. The mere fact that the power of statistics increased by adding the information from the wild group indicates that a directional pattern of selection is being followed in the wild. This clearly suggests that these variations in EEB are attributed to their wild caught parents, in accordance with the earlier statement by Drent and coworkers (2003).

However, considering the smaller sample size used in this study, it is difficult to come to a definite conclusion. The effects of other possible candidate genes involved in this phenotype also need to be addressed. Moreover, the genotype environment interaction should also be taken in to consideration. Hence, a study involving more individuals from both the selection lines and wild together with other candidate genes could provide further support to these findings.

5 Conclusion

Based on the haplotype information, selected 6 SNPs representing these blocks could be typed for genotyping a larger population. A study involving more individuals, both from the captive and wild population could throw more light in understanding the variations in these two groups and also to explain natural selection and evolution. Further, a detailed scanning of the introns 3 regions found in this study might be very interesting to find any regulatory mechanism in this region. Thus, the role of dopamine receptor gene as well as other candidate genes involved in the personality traits could be explained.

6 References

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)*, 21(2), 263-265.

Berlin, S., Quintela, M. & Höglund, J., 2008. A multilocus assay reveals high nucleotide diversity and limited differentiation among Scandinavian willow grouse (Lagopus lagopus). *BMC Genetics*, 9, 89.

Brudno, M., Gelfand, M. S., Spengler, S., Zorn, M., Dubchak, I., & Conboy, J. G. (2001). Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Research*, 29(11), 2338-2348.

Dietrich, R. C., Shukla, G. C., Fuller, J. D., & Padgett, R. A. (2001). Alternative splicing of U12-dependent introns in vivo responds to purine-rich enhancers. *RNA*, 7(10), 1378–1388.

Drent, P.J., van Oers, K. & van Noordwijk, A.J., 2003. Realized heritability of personalities in the great tit (Parus major). *Proceedings. Biological Sciences / The Royal Society*, 270(1510), 45-51.

Fidler, A.E ., Oers van, K ., Drent, P. J., Kuhn, S., Mueller, J .C and Kempenaers, B.(2007) Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proc. R. Soc. B.* 274, 1685–1691

Gelernter, J., Kranzler, H., Coccaro, E., Siever, L., New, A., & Mulgrew, C. L. (1997). D4 dopamine-receptor (DRD4) alleles and novelty seeking in substance-dependent, personality-disorder, and control subjects. *American Journal of Human Genetics*, 61(5), 1144-1152.

Groothuis, T.G. & Carere, C., 2005. Avian personalities: characterization and epigenesis. *Neuroscience & Biobehavioral Reviews*, 29(1), 137-150.

Kluger, A.N., Siegfried, Z. & Ebstein, R.P., 2002. A meta-analysis of the association between DRD4 polymorphism and novelty seeking. *Molecular Psychiatry*, 7(7), 712-717.

Mitsuyasu, H. et al., 2001. Association analysis of polymorphisms in the upstream region of the human dopamine D4 receptor gene (DRD4) with schizophrenia and personality traits. *Journal of Human Genetics*, 46(1), 26-31.

Paterson, A.D., Sunohara, G.A. & Kennedy, J.L., 1999. Dopamine D4 receptor gene: novelty or nonsense? *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 21(1), 3-16.

REFERENCES

Rang, H, P., Dale, M.M and Ritter, J.M. (2001) Other transmitters and modulators.Pharmacology, *4th edition. Edinbur*gh, UK: Harcourt Publishers Ltd*:483–499.*

Sanak, M., Zelek-Molik, A., Nalepa, I., Wegrzyn, J., & Wciorka, J. (2005). The dopamine D4 receptor VNTR in Polish schizophrenia patients. *Schizophrenia Research, 73*(1), 129-131.

Stephens, M., Smith, N.J. & Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68(4), 978-989.

Stephens, M. & Scheet, P., 2005. Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. *The American Journal of Human Genetics*, 76(3), 449-462.

Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386

Reale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, 82(2), 291-318.

van Oers, K., de Jong, G., van Noordwijk, A. J., Kempenaers, B., & Drent, P. J. (2005). Contribution of genetics to the study of animal personalities: a review of case studies. *Behaviour* 142, 1191-1212.

Wang, Z., & Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA (New York, N.Y.)*, 14(5), 802-813.

Appendix A

Details of the primers used and region selected in the study

Product NO	Product SIZE	Primer	Remarks
1	**632bp**	**FP: 5-GGCTCAGTGAAAGTGGTTCC-3** **RP: 5-CAGCAGCTGATCCACACAAT-3**	**Sequence selected between position 15-646**
2	**466bp**	**FP: 5-GGTAACCATGACCCTTTCCA-3** **RP: 5-GGGCTGAGGTGTCTTACTGC-3**	**Sequence selected between position 737-1202**
3	**501bp**	**FP: 5-TTGCTTGGCAGGTTGTTGAT-3** **RP: 5-AGGCCAAGGTAGAGACCATTC-3**	**Sequence selected between position 2618-3118**
4	**526bp**	**FP: 5-TGGGCAGAAGGCACTTATCT-3** **RP: 5- GGGATGCCTCCACTTAATGA-3**	**Sequence selected between position 4880-5405**
5	**675bp**	**FP: 5-AATCACCAAGGATGGCAGAG-3** **RP: 5-GATCCCTGGTGTCAGCAGAT-3**	**Sequence selected between position 8555- 9229**
6	**527bp**	**FP: 5-CCTTGATGGAGAGAGAGCAGA-3** **RP: 5-AGCTCAAGCACTCAGGGAAA-3**	**Sequence selected between position 10022-10548**
7	**489bp**	**FP: 5-AGGAGGGGGTACAAAACCAC-3** **RP: 5-ACTGCATGGAAGGGAAAAAT-3**	**Indel 1 (15bp) Region 713-727**
8	**300bp**	**FP:5- CTGCAGCCTCCTGGAATTAG-3** **RP: 5-TCTCAGCTGCAGCACCTTT-3**	**Indel 3 (48bp) Region 7054-7101**
9	**413bp**	**FP:5-CCCAAAGGATGGTGGAATTT-3** **RP:5-CTGCCATCCTTGGTGATTTT-3**	**Indel 5 (12p) Region 8489-8500**

APPENDIX A

10	**502bp**	**FP: 5-ACCAGAGCAGTGCCAAAAAC-3** **RP:5-CTTTGCGGAAGAAGTTCCTG-3**	**Indel 7 (4bp)** **Region 10622-10625**
11	**861bp**	**FP: 5-TCAGTCCCCAGGTCTCTCTG-3** **RP: 5-GAGTGCAAGCTGGAACCAAG-3**	**Intron 1 region** **3389-4230**
12	**851bp**	**FP: 5-CACATGTGGACTGTGCTGTG-3** **RP: 5-TGTCACAGCCCCCAGAATAC-3**	**Intron 1 region** **5401-5286**

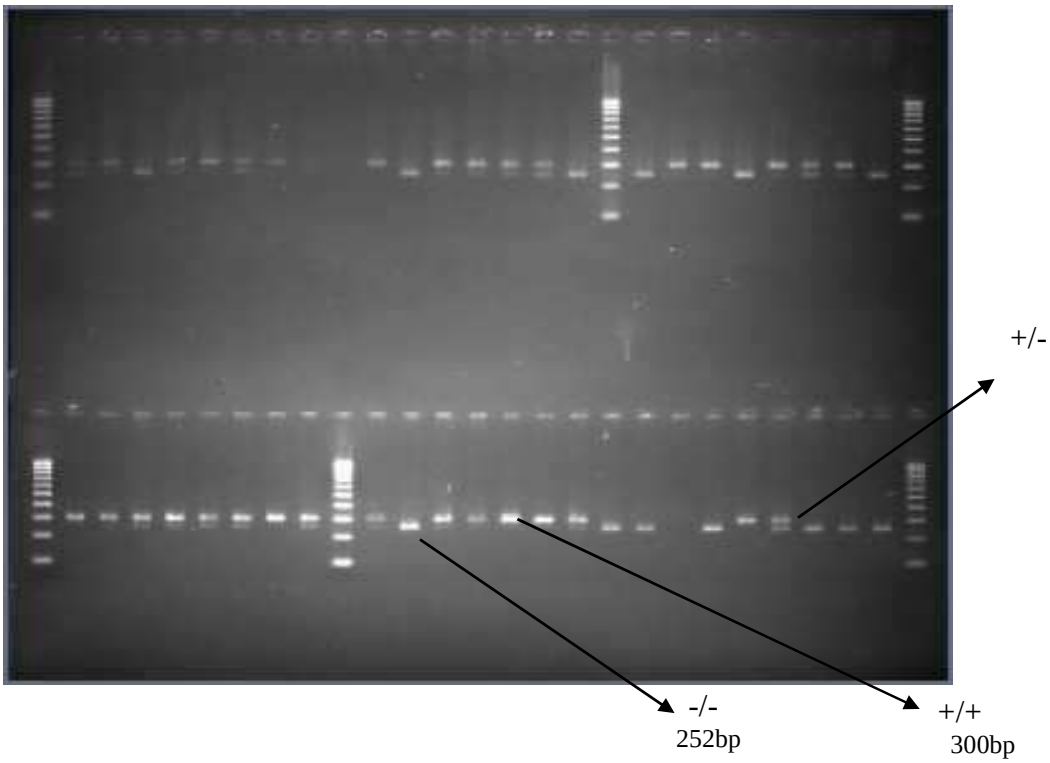
Appendix B

PCR conditions used

Product No.	**Step lengths (s)**		**No. of cycles**	**Annealing Temp (T_a) in ^oC**
	Annealing	**Extension**		
1	Hotstart*	Hotstart*	Hotstart*	Hotstart*
2	45	90	35	55
3	45	90	35	55
4	45	90	35	55
5	45	90	35	55
6	45	90	35	55
7	45	90	35	58
8	45	90	35	55
9	45	90	35	55
10	45	90	35	55
11	45	90	35	60
12	45	90	35	55

*Hot start PCR profile consisted of the following time temperature combinations: 96 ^{o}C ,15.0 min (1x), (95 ^{o}C, 0.45sec ; 63 ^{o}C, 0.45 sec; 72 ^{o}C, 30 sec) 5x, (95 ^{o}C, 0.45sec; 61 ^{o}C, 0.45 sec, 72 ^{o}C, 0.30 sec)32x and 72 ^{o}C for 7 min .

Appendix C



Gel picture showing various genotypes of indel 3, separated by Agar Gel electrophoresis of the PCR product

Appendix D

Complete information about SNPs and indels newly found as well as previously known and validated by this study are given

```
  1 ctcagaataa aaagGGCTCA GTGAAAGTGG TTCCgctgtc ccaacccctc actgagtgaa
                    ←  FP product 1      ←
 61 ggctgtgcca ggagtcaggc tttagggatc aaagccagat aggaaaagct gtgaccagat
121 gatgGaaacc atcttccttg tctgaacttc ccagcagatg aggtatcact gagctctgaa
        A 125
181 Tctggcagat gtgcaaccat gctTgtgtcc tccTgagtcc aacacaaaac cattccCaaa
    C 181                    A 204      A 214                236 G
241 aggaatgcag agatacaagt atgtgcatac ccactccagt gctggtgaca gacagatcta
301 gagcaagacc tggtcccact caggccaggt ggaattgtct ctcctgtctt tgcacagtaC
                                                                360 G
361 aagagggaga tgactggaaa aacaattcca tgggtattgg gatcaacatg agaatgggaa
421 accttccctg ggatgggagA GGAGGGGGTA CAAAACCACa gttactaaag acatggtatt
                        ← Indel 1 FP(product7)←
481 cttgctggct ttagaaatTa cctGgaagcc ccacctctgg aagcagaatt tgaggacaac
                    499 G     A 504
541 cagactgttg tccaagtgct taaaccaagg gaattttcTc ctactcGtgt atgaaattGc
                                          579 C    587 A        599 A
601 agagccacaG cagctgggca aaacagATTG TGTGGATCAG CTGCTGttag gccagtccaa
          610 T                ←   RP product 1      ←
661 ggGaaggaca gtgcttggat ctgtgtcctG tggctgacaC cagggctgtg gcctttccat
      A 663                         690 A      T 700   (713-727 indel-
721 gctgcacatg ctgaagGGTA ACCATGACCC TTTCCAaaaa aagagtaagg gaactttggg
    Polymorphism(1)  ←  FP   product2)    ←
781 gcaGcaaggg cagggtttgg agagaggatt tctcccacaa gtcttaatgG tttgTatgaa
       A 784                                        830 T 835 G
841 actctgatag ctgtagccta taacagagct tagaaTcaat ttgtctcata aaaagcttcc
                                        876 G
901 aatATAGTAT TTTTCCCTTC CATgcagtat aaatgtcact ttattgctat gcCttctctt
     ← Indel 1 RP(product7)←  TATA_signal (928-933)     953 T
961 cccactgcct tccacaaaca TtagGctgac ctgaatgctt acctgGcatt cacatctggg
                      981 C   A 985            1006 A
1021 ctgaggtttg gtggcacagc tgcagagatg cccacagagt ctgtatcagc tccccccatct
1081 tccaagtcat ctgggctaga aatgggaggt cacatctccc tgtgcttcct gatcctgcag
```

APPENDIX D

1141 gtgcataggt acccctcaaa cagcattttt ctCtcatttt ctGCAGTAAG ACACCTCAGC

1173 G ←RP product2 C1196

1201 CCccacgttc agccaccct catgacctgC caccccatct ttgctcccac acttgctgcc

RP2← T 1230

1213-1218 Indel Polymorphism(2)

1261 tatctggttt cttcctgtat cctacatttc tgcagttgtt attcctcccc aagcatgggt

1321 aaggtgcaag cttccttttg aagtgcatcc tgcatttgcc aacttatcta gagcactttg

1381 aatccctact ctgcgctaca agatgctcgc agcatcctgt cttggcacta tcttcaaatt

1441 tcattatgga ctgtctggat cttcactcag tttaattaca aattgctaaa cattaatgga

1501 ttctacacac acatcggaat acaatttccc agttattcac ccgatccgta ctgattccaa

1561 tttctttcct ctaacattgc ttaaagaaca gccccatgcg agaacacagc gagaccatgt

1621 gagaacgcag caagacctct gcagcgctcg ggacaccggg acaccgggaa cctgtccccc

1681 ccgcatcggc cacccttaa gccgcttggt gacccagttg tccccgcggg gcccggccgg

1741 ggccgggcgg ggcggcggcg gaggctcctc ccgcggctcg gcgggcagct cccggcggcc

1801 ggCccggctc cgtgcgcggg gctgcggggg cagcgccgcg gggccatggg caacggcacc

T SNP (Exon 1)

1861 gccggacccc cgcccgccgg agccggccac agcatcgccg ccctggtgct cggcatcctc

1921 ctcatcctcc tcatcgtcgg cggcaacggg ctcgtctgtc tgagcgtctg cacggagcgg

1981 gcgctcaaga ccaccaccaa ctacttcatc gtcagcctcg ccgtggccga cctgctgctc

2041 gccctcctcg tcctgccct ctacgtctac tccgaggtga gacagcccgg ggccaccggg

2101 gcacggtggg ctggggcttg tgtgcccgcc gagcctgcag gttacgaggc agcgggggg

2161 caagagacca agtgccgagg gtcaaagccg ctctccgggc ttgctcccca ccgcccgcgg

2221 gcccgggaca gccctgtgcc cgggatgtgg cggtcagacc ctccggctct tacctgcgag

2281 cctcacggca tcgcacagcg ctgcgggacc gcgggatgcg gggagcggag ccggctccgt

2341 gcgggttcct gggatccccg tccgaacgga gatgccccga gtggaagcga agcggAgtgg

2396 G

2401 ctgtcccctg gcttgcacgc tggccaggtt accgtcagca aatggatagg aagagacagg

2461 tgctgtgctc tgtttggcat aaagtgagca gatctcggtg ataaaaggag atctttagtt

2521 tctttagcca ggaaacaact ttccgtgttg gctgagaagg agaagccgtc tttcccggga

2581 gctggatggg atttggtcac cctgcaccga ggtggctTTG CTTGGCAGGT TGTTGATtgg

← FP product 3 ←

2641 cacaggcctg acccagcata tgatggcagt cgtggagagc cagctcaggg acatggggca

2701 gagttgcttg ttcactcccC aaacctgttc acccccatga Ctgtggcagc tgagggaaag

2720 T 2741 T

2761 aacaacagaC actcccAagg ctgagtggtt ctctctgtaa catttactgc cagggaggaa

2770 T 2777 G

2821 taaaccttg tgcttgcagg gggatgttta aaacctgcta gtaccagCTg gtggctcttc

2868 TG 2869

APPENDIX D

2881 cctgctcaag aggatggaaa catccaCtgg gccaaatagc Aggtttttct aCccatcctg

2907 A 2921 G G 2932

2941 gtggtagctg gtagtgtgga tcttcttccc agctgtggaa gagcAtctcc cctcttgtcc

2985 C

3001 tctctggcaa gccatgcaga gcaccccatg gcccagcagg ccAtggccat gcaggcgtgt

(new indel)3031 G 3043

3061 gctcccctaa gcaaggcagg ctgcagatta gagcttaGAA TGGTCTCTAC CTTGGCCTag

← RP product3 ←

3121 gattttgcag gacattgtga cctctctgga gacaggtatt gtgttcaata cctgcaggtt

3181 catctgggga cccatgagaa gaaccaatga agaactggag cattaggtcc catcttttc

3241 cactgtgctc ttggttttgg tggatttgct tttccctcat tacctcaatt ttgctggtta

3301 gaaaggcaca ttcatcattc ttctcagcta ctttttgcta attgggatcc cttttcttcc

3361 ttccactttc tgttgtctct cgatgcctTC AGTCCCCAGG TCTCTCTGtg ggcagtcatc

← FP product 11 ←

3421 aacaccagag tggtttcctt ttgccttcag gtaggagttt tgctggaGta caTaaaatac

3468 A 3473G

3481 agattaacGt attctggcgt acctctgaaa atgcttccaa tggggcagga gtattgtggg

3489 A

3541 gctttccttg aataccacag acattccaca agttctggat tctcaggtaa caggtgttca

3601 tacctgtatg agctttaaaa actggggttg tccagcttgt ggtgacactt gatatattca

3661 cttccatggc tgtttataaa catcagttct gtgagggaat tcttccatgt ctctgctccc

3721 aggatattaa acatgggcag tgaaggtgct gtaagacaga agtctttcct tctcatcctT

3780 C

3781 tgcagctcct gcagccaggc aggcattcaG atGcacaaaa cagcacagga tttgcactcc

3810 T A 3813

3841 tggcactcgg agggaatgca tggagcaggt gtgggtagtg ctgaacagag agagagagca

3901 cggtgcagca gtcacagccc cctcaagcag gcagtgggtc tctgatgtgt agagcaactc

3961 cagcgctgca gcagttctca ttccatgcct ggaagcttct ctgtaggCtc agagcctgca

T 4008

4021 agtgttcatt ttaattgtCt cCtactgatg caggatttcc agctcatttt gttctgaaat

4039 T G 4042

4081 aaggaaaagg gagagtaacc actgaatcca tgccagagga cagaattttt gtggtttctt

4141 ttctttgtct ttttttcTtt ggagGcaaaa ctcagaataa gatgttggga tgcaagatgc

4158 C 4165 T

4201 agggtgtcct gattgctctt tagaactagC TTGGTTCCAG CTTGCACTCt ttaccaaaca

← RP product 11 ←

4261 tggtagccac tgaaaatatg gtctggacag tcctgaatgt tcacctcaac attcattctt

4321 ggtgtgggga ttttgttttg ttgttttaaa ctttggtaca actcctgaca cagtaaaggt

APPENDIX D

```
4381 ggattctggg ggcctattgg aaaactcaaa aaacagctga ggtggtgggt ttgtgctttt
4441 tttagcagga acagagcaat aaactgagca ttttttcatt caccttgtgc tgttttcctg
4501 tgaggtggca tcttgtgttc cagttttgga aacattcctc ctaatcagag ctttagaaga
4561 cagcaggagg gacttagctt ttgaggtgac atctgacatc ctCacaaaca acagatctcc
                                              4603 T
4621 tccaagcccc agtggggatg ggattccttc ctgtgggatg tggatgatcc tttgctgcca
4681 ttccagaaac tggagcacag tgggagagtc cttgatattg gcttttattt ctggttgtta
4741 agaaagctca tttcagacac tggcagcaat acatactctc aaaacacata ctctctttca
4801 actggactca tccaaatatg gagcaaagac atgttccaat gttttctagt tgtaattacc
4861 acacagctgc ctgccctgaT GGGCAGAAGG CACTTATCTt cctgcttgct tgggaatctc
                          ←  FP  Product 4     ←
4921 agatcttagc agtatgatgg ggagaggcaa tggtaacttC catggggaag ttgcaCtgct
                                          4960 A            4976 T
4981 cttctgcCag ctttggtttt ctgatgggtg tgtcaagtgc ttttcagcaa tcaatccagt
           T 4988
5041 gccagagagg cagagaagca aggggaaagg agatggcctg gggcactccc tggcattaac
5101 tgttactcca ggagcagagt tctgctttga catgaagaga aaccaggaga agTtcccaga
                                                    5153 A
5161 gagtgcctgg agcattggTa gcactgaaac agggggctct gaggaaggag agctatGgtg
                  5179 A                                  5217 A
5221 cttgtggggt tactgaagga tatcacttgg ctctgtgtgg ctttgggagg gagctgcaTg
                                                          5279 C
5281 gGtccagagg ctcgccttgt gtgtaactgc ttcctcttcc tagagactgt gacagtggca
      C 5282
5341 ggagtcccaa agactctgga acacttgtag gagtttatac agcccTCATT AAGTGGAGGC
                                              ← RP product4
5401 ATCCCagcaa Ctggggcagg caccaggaca agctcCACAT GTGGACTGTG CTGTGgagca
       RP4←     T 5411                  ←      FP Product 12    ←
5461 cacctgtacc tgtacctgct gctcctgtga tcagctctCt gctaatccct gccGagcagc
                                          5499 T                A 5514
5521 agagctgcag aAcaggcctg agccaggaag ctgctggCag gagctgaggg tgttgcttat
                 G 5532                    A 5558
5581 gtaaacatct ttgccactaa gacctgtgta ggtgattatg gaaaagagga aacagtttga
5641 gcagatttgc atGtttacct ggctccttct ccaggtcagg gaagatgagt gttcccagAa
             5653 A                                          5699 C
5701 gcCaataaaa accacatgag ctatcttctt atcttggtga gggtgatgag ttctgcctaa
       T 5703
```

APPENDIX D

5761 aactttcagg gtatcactgg tttggtatca gtacatctat tttcttgcaG tattttctgt

A 5810

5821 gtgcctgtga taTagttcct ctaatgagag gccagacctc taccctctcc aaattgccct

5833 C

5881 ggctctctgt aaatgcgtat cagcttgttt gcagtaggat ttgggattag tctttcctag

5941 aatgttacag ttctagtgga aatagatgtt attctgggtc tggctgatgt ctgagccaca

6001 caaagtaccg ggaaggtgag acaggccGgc aattactaac cagtgcttag ggtttggtgt

A 6028

6061 tttttcagat accagtaact ggggaggggt gtcttggaat tcttggtcct cttaatgtct

6121 caggtggatg atcaggcttg ctcatgatct ctgaaaaact ttgacctcaa tgatttaagt

6181 taaAaaacaa catcagagca gacatgaggc acctctgttt tatgtcccac tccctcttct

G 6184

6241 tcctctcccc ttgcaggctc cattgtGTAT TCTGGGGGCT GTGACAgttg tcacaactta

← RP Product 12 ←

6301 tgacaaaaag ccccaagtca ggttctggtg gacatggggc caattctttg actctgatgc

6361 tggactttac ccactggcag ggactgttcc tcttccagct catgtcacag acttctcact

6421 ccatgcctca catatctttg agccatgGct gctctgaggc aggtgcttgt tcaaagtcca

6448 A

6481 gctgggcatc tccgaattcc ctttcaggct tctgctgctc cctgcatgtg gcaggactga

6541 gacaagagct gatatcagag gaaataggac taaaaccttc ggttctgtcc tggctcatgc

6601 cctcaggagc tctgctggca tcaggctggt ttagaaggca gagttttcct tacaaatcag

6661 ctgtgcttag gaataaggat accagaatgg gctgacttag aaggcagttg caagggtgtt

6721 taagaagcca agtctAtctc ctgccatcca ccatcctata aaaatttctg cattgtattc

6736 G

6781 tggaaaagtg gagaacccaa agagccgctg aaggtgcagt tgctctggag cacttcttaa

6841 agaaagcttg agctgtttgt gatccataaa tctttacact gtgtaccttg atgggctaat

6901 ctgCTGCAGC CTCCTGGAAT TAGgtactcc ttaatcacat ggcccagcaa agcctcttat

← Indel3 FP (product 8)←

6961 ccaaagttAc tctcccatta gaggggtcct tggcatgttt tgccaaggtg ggcaggtgac

6969 C

7021 aaacacccta gaatgggaag ttctctccta gcaAGGTGGG CAGGTGACAA ACACCCTAGA

7006 - 7101 repeat_region

7081 ATGGGAAGTT CTCTCCTAGC Accttccca aagccagagt cttgagggtt ggtggggaaa

7054 – 7101 Indel Polymorphism (3)

7141 gggaagggtt tatagcattt tccaacttta atctcacatg atccAAAGGT GCTGCAGCTG

←Indel3 RP(product 8)

7201 AGAatccttg tccttttcag aatcaattag ttcccccaaa tcacaaatag gtttggaaaa

RP←

APPENDIX D

```
7261 ctgcagccca aaatcaaatt ttctctctca taggctgaat tgaacctgtc aagatcccag
7321 tgagcttggc agaaatttgt tttttgatga ttcaggtcca aattcagttt ggagtggctg
7381 cattgtgcta gagaagaggc tgaatgaagc cctaaggtct gactcgagct cccctggagg
7441 tatctcaaga agatgctgta caaaaatagc aaaacagtta cttcagagac tcctttggct
7501 ttacagctcc tctaatcata atcctgtgct tgacatcatt actggctgat attaatccct
7561 gcagagctgc ctgccctgtt gctggcaggg gctgggggtt cagcactttc ccggtgctct
7621 gtggagcgtg ttccccacgt ggctgcactc tgagctggga tgaatgaaac cagaacctgt
7681 tgtttagcac cagtacagga aaaaattcat acctgcttcc tgaaacctgc aacatgaaat
7741 ggcagcaaca ttagactcta tgagcactgg ctggctattg ctatagagaa aaaagtccct

7801 ggaatttata tatatatata tctaaataat tttaacacag cacatgttct aaagataaaa
7861 atatgtaata aattatatat atattattaga tgattcacat tatctgtaga taaaaacaca
                    7884-7885 Indel Polymorphism (4)
7921 aagaaaataa aatatattac atgtagcaac atatatacac agtagtattt atcatacata
7981 ttttaataaa agcatataga gagtacagat aatgtttggt attttgcaat ataaaaataa
8041 agtgtttgag tgagacatga caaggggcca ttgctactga acacttttt ttcaagctaa
8101 aatcatactt atttgtacat tttattgggt ttttaactga acctctctct gcagagggaC
                                                                    ←
8161 CCAAAGGATG GTGGAATTTt cctctctctc aaactcagag gaagaattcc ctactgtcct
     Indel 5 FP(product9)←
8221 tgtggtgaca gcctattgag cttcctggtt tacctgtgca ggtgcaagca aacagataag
8281 tctctGcctg tgctcaagga gttcAtggga ataatggcac aggGatgaga tTgaggtgac
     8286 A               8305 G               8324 A        C 8332
8341 ctatttctgt gtcagcctaa aaagtgcttt aatgttaaaa gaacaaGctg ggtttgtacc
                                                  8387 A
8401 tgggtgagat gcactcGact agaacaaagc Actgcaagta gtttctagct gtaccagcat
                 8417 A          8431 G
8461 cagccattgc tgtcctgctg tctgtcctgc tgtctgtcct ctgctatagg ctgtgcagag
                        8489 -8500 Indel Polymorphism (5)
8521 cctgtgcttc tggctgagac atgactgaac tgAAAATCAC CAAGGATGGC AGAGccaaca
                                      ←   FP product 5     ←
                                   ←Indel5 RP product 9  ←
8581 tttccaggtg gttcagagta acaccggtcc ctgctgctct gccctggtgc ctgtgaccga
8641 ggggagagaa tcagaggttg ttactcactg cctggGcctt gtatccagag gctggacttg
                                      8676 A
8701 tgccctgcca ggatggaaca gtggtgtaac acagggaaat cattcacatc Gtctgcacac
                                                  8751 A
```

APPENDIX D

8761 catgggttgg agcagttggg acacaggaat Agaagagaag ctttgtgctc tgtctcttcc

8791 G

8821 atgcttgggg actgagctta ggatgttgtg cagtcagtac tgccaagctg gggagatcct

8881 ggggtcctga acgggggggg tttgaggagt ttggaggtac ctgccctcac ccaaggcagc

8941 acaatgtcct gtgtccattt ctctgcccag catgactcac ttggctgatg gtgagtctga

9001 tggtctgaaa gctcagttct ccagatgggc cagaaccaca agccCtgagc agGctgtggg

9045 T A 9053

9061 ctctgtttgg ggatgattcc catgtgctgc tgggagcatt gtgccatagg aggAACACTA

9121 TGGCAGCTTT GGGctgcttg tttggaggaa acatgCgtgg ctgaggcacg Taggcatggc

9156 A 9171 C

9181 catgtccctg tcattttaaa ggacacagcA TCTGCTGACA CCAGGGATCt tttcagttcc

← RP product 5 ←

9241 agggaggagt gtggtccctc agcacggtgc tgtgcgatgc cctgatgacc atggacgtga

(Exon 2)

9301 tgctgtgcac agcctccatc ttcaacctgt gtgctatcag cgtggatcgg tgagtggctt

9361 cctgctctgg ctgtgcctgg gcagcacgtt ggttccatag gcccttgcca gtgtcccaag

9421 cagggattct acccccccctc aggagggtgt tctgccctaa tgtccttcac cgctgggtgt

9481 gtgtctcacc agtggtggcc aagcccgttg tgagatgaac cagagctgtc ccttggctgt

9541 tctggttgta gcccaggcct cgactgtttg taaattaatg aggggaagga aggacagtga

9591-9594 Indel Polymorphism(6)

9601 caaggagctt ggggcccctc agttacatta ccagtggtgc caccaccaca ccaggactga

9661 ctcatccagg gctgtgttgg caggttcatc gctgttcaaa tcccgctcaa ctacaaccgg

9721 cgacagatcg acctacggca gctgatcctt atatccacca cctggatatt cgcctttgct

9781 gtggcttccc cagtcatatt tggtctcaac aatgtcccaa accgggaccc cagcttgtgc

9841 caattggagg atgacaacta catcgtgtat tcctccatct gctccttctt catcccatgc

9901 cctgtcatgc tggtgctgta ctgtggcatg ttccaaggac tcaagcgctg ggaagaagcc

9961 cggaaggcca agctgagagg ctgcatctat ggagccaaca ggaagctgta tcacccccca

10021 aCCTTGATGG AGAGAGAGCA GAcccggctg gggctgctgg actgcagcag cccctatgcc

← FP product 6 ←

10081 cgtgcCggcc tccctgggga gtgtgggatg aacagtggga tccagactgt gtcctaccca

T 10086 (SNP 830) (Exon 3)

10141 cacctcaggt acccgcaccc agggcacggg cacaagcggg ccaagatcaa cggccgggag

10201 cgcaaggcca tgcgGgtgct gccggtcgtc gtcggtgagt ggctgtcagg ggtggctggg

A 10215

10261 gagggtggga aatgtgccag cttgtcccac cacagcgctc agactggcag gAcctgggtg

10312 G

APPENDIX D

10321 aaACCAGAGC AGTGCCAAAA ACcAccaccc ttgggaaagg tttataaccc aacctccctt

←Indel 7 FP product10 ← G 10344

10381 caaggcaaac Ttgtgactcc Caatatctta agttcaaagc caaagcaaga cttacttaag

(SNP79)C 10391 G 10401

10441 acttttcatc agtgtttagg ataggagaga catcccagtt tttgtctgAg cataagccca

10489 G(SNP 81)

10501 gctccacttg tcagcacaac **tgctgcctTT TCCCTGAGTG CTTGAGCTgt** gtatatttat

← RP product 6 ←

T 10533

10561 ttttttttg gctGaggctt Cgttaaatta aacccttctg aacattcaca tcaccttctc

10574 C T 10581

10621 tctttattgt cctcctggca ggtgctttcc tcttctgctg gacacctttt tttgtggtgc

10622-10625 Indel Polymorphism(7)

10681 acattaccag ggctctctgc aagtcctgct ccatcccccc tcaagtcacc agcactgtca

10741 cttggctggg ctacgtcaac agtgctctca accccatcat ttacaccgtg ttcaacgccg

10801 agtt**CAGGAA CTTCTTCCGC AAAG**tcttgc atgtcttctg ctgagccctc tgcacaggag

←Indel 7 RP product 10←

10861 gaaccaccgg gcaggaggaa ccactgggtc atttttt (Exon 4)

*Areas highlighted in yellow are the sequenced regions. SNPs shown in blue in these areas are the ones which are validated in this study. Those shown in **pink** are the newly found SNPs. SNPs shown in **black text** are the ones missed because either they were absent or couldn't be typed. Indels are shown in **green**. All information about the primer designed is also shown in the text. Also, all other details known about the gene from the previous study by Fidler *et al*, 2007 is also given here.

Abbreviations used:

FP- Forward Primer

RP- Reverse Primer

Appendix E

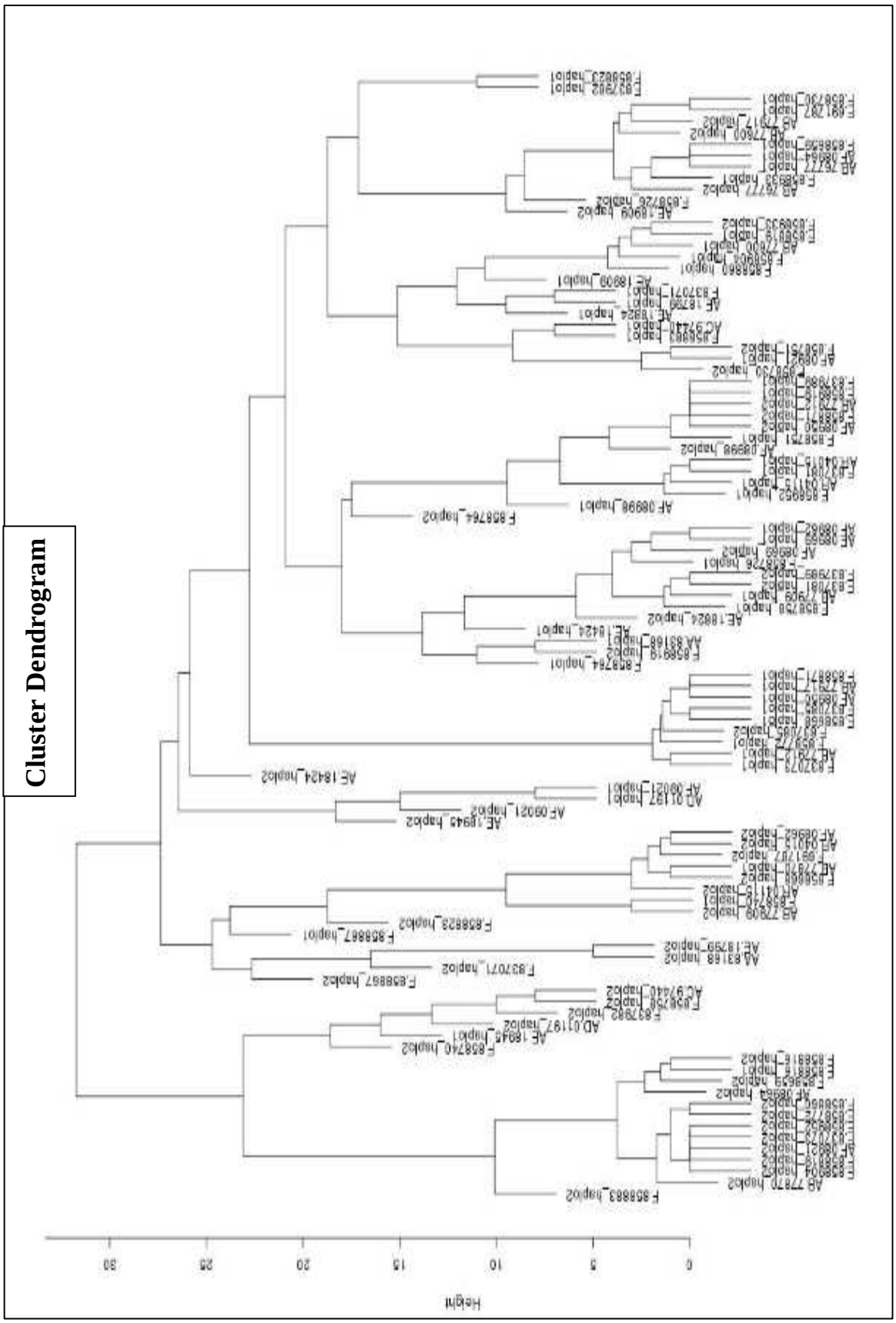
List of haplotypes found in best reconstruction using Phase, with counts

1	GTTTCCTGTGGGGGCAGTTCTGGCCCACGCACAAATGACCCCGATCATGCGCGACACCGCGBGGGTGGAAGGACGCGATCACGCA	2.000000
2	GTTTCCTGTGGGGGCAGTTCTGGCCCACGCACAAATGATCCCGATCATGCGCGACACCGCGAGGGTGGAAGGACGCGATCACGCA	1.000000
3	GTTTCCTGTGGGGGCAGTTCTGGCCCACGCACAAATGATCCCGATCATGCGCGACACCGCGBGGGTGGAAGGACGCGATCACGCA	3.000000
4	GTTTCCTGTGGGGGCAGTTCTGGCCCACGCACAAATGATCGCGATCATGCGCGACACCGCGAGGGTGGAAGGACGCGATCACGCA	1.000000
5	GTTTCCTGTGGGGGCAGTTCTGGCCCACGCACAAATGATCGCGATCATGCGCGACACCGCGBGGGTGGAAGGACGCGATCACGCA	1.000000
6	GCTTCGTGTGGGGGCAGTTCTGGGTTGCTAGCABACGGCCCCGATCATGTCCGACGCCACABGAGTGGGBGGACGCGATCACGCA	1.000000
7	GCTTCGTGTGGGGGCAGTTCTGGGTTGTGCGCCAACGACCCCGATCATGCGCGACGCCGCGAAAGCGAAAGGGCGCGATCACGCA	1.000000
8	GCTTCGTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGCCCCGATCATGCGCGACGACGCABAAGCGAAAGGGCGCGATCACGCA	1.000000
9	GCTTCGTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGCCCCTATCATGCGCGACAACGCGBGAGCGAAAAGGCGTGACGGCCTA	1.000000
10	GCTTCGTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGTCCCTATCATGCGCGACACCGCGBAAGCGAAAGGGCGTGGCGGCCTA	1.000000
11	GCTTCGTGTGGGGGCAGTTCTGGCTTGCTAGCAAGCGGCCCCGATCATGTCCGACGCCACABGAGTGGGBGGACGCGATCACGCA	2.000000
12	GCTTCGTGTGGGGGCAGTTCTGGCTTGCTAGCABACGGCCCCGATCATGTCCGACGCCACABGAGTGGGAGGACGCGATCACGCB	3.000000
13	GCTTCGTGTGGGGGCAGTTCTGGCTTGCTAGCABACTGCCCCGATCATGTCCGACGCCACABGAGTGGGAGGACGTGATCACGCA	1.000000
14	GCTTCGTGTGGGGGCAGGTCTGGCTTGCGCGCCAACGGCCCCTATCATGCGCGACACCGCGBAAGCGAAAAGGCGTGACGGCCTA	1.000000
15	GCTTCGTGTGGGGGCAGGTCTGGCTTGCTAGCABACGGCCCCGATCATGTCCGACGCCACABGAGTGGGAGGACGCGATCACGCA	1.000000
16	GCTTCGTGTGGGGGCAGGTCTGGCTTGTGCGCCAATGGTCGCGCCCTTGTGCGACGACGTGAGAGTGGGAGGACGCGATCACGCB	1.000000
17	GCTTCGTGTGGGGGCATGTCTGGCTTGCTAGCABATGGCTCCTATCATGCGCGACAATGCGBAAATGGGAAAGTGTGACGGCCTA	1.000000
18	GCTTCGTGTGGGGGTAGTTCTGGGTTGCGCGCCAACGGCCCCTATCTTGTGCGACGACGTGAGAGTGGGBGGACGCGATCACGCB	1.000000
19	GCTTCGTGTGGGGGTAGTTCTGGGTTGCGCGCCAACGGCCCTGCCCATGCGCGACGACGTGAGAGTGAAAAGGCGCGATCACGCA	1.000000
20	GCTTCGTGTGGGGGTAGTTCTGGGTTGTGCGCCAACGGCCCTGCCCATGCGCGACGACGTGAGAGTGGGAGGACGTGACGGCCTA	1.000000
21	GCTTCGTGTGGGGACAGTTCTGGGTTGCGCGCCAACGGCCCCGATCATATCTAAAGCCGCGBAAGCGAAAGGGCGCGATCACCCA	1.000000
22	GCTTCGTGTGGGGACAGTTCTGGGTTGTGCGCCAACGGCCCCGATCATGCGCGACGCCGCGAAAGCGAAAGGGCGCGATCACGCA	1.000000
23	GCTTCGTGTGGGGACAGTTCCAGCTTGCTCGCABACGGCCCCGATCATATCTAGAGCCGCGBAAGCGAAAGGGCGCGATCACCTA	1.000000
24	GCTTCGTGTGGGGACBTTGTCAGGTTGCGCGCCAACGGTCCCGATCATGCGCGACACCGCGBAAATGGGAGGACGTGACGGCCTA	1.000000
25	GCTTCGTGTGGGGATAGTTCTGGGTTGCGCGCCAACGGCCCCGATCATGCGCGACACCGCGBGAGTGGGAGGACGTGATCACGCA	1.000000
26	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAACGGCCCCGATCATGCGCGACGCCGCGAAAGTGAAAAGGCGCGATCACGCA	1.000000
27	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAACGGCCCCGATCATGCGCGACGCCGCGAGAGTGGGAGGACGTGACGGCCTA	1.000000
28	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAACGGCCCCGATCATGCGCGACGCCGCGBGAGTGGGAGGACGTGACGGCCTA	2.000000
29	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAACGGCCCCGATCATGCGCGACGCCGCGBGAGTGGGAGGACATGACGGCCTA	1.000000
30	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAACGGCCCTGCCCATGCGCGACGACGTGAGAGTGGGAGGACGCGATCACGCB	1.000000
31	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAACGGCCCTGCCCATGCGCGACGACGTGAGAGTGGGAGGACGTGACGGCCTA	5.000000
32	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAACGGTCCCGATCATGCGCGACGCCGCGAAAGCGAAAGGGCGCGATCACGCA	1.000000
33	GCTTCGTGTGGGGATAGTTCTGGGTTGTGCGCCAATGGTCGCGATCATGCGCGACGCCGCGAAAGCGAAAGGGCGCGATCACGCA	1.000000
34	GCTTCGTGTGGGGATAGTTCTGGGCCACTCACAAACGGCCCTGCCCATGCGCGACGACGTGAGAGTGGGAGGACGTGACGGCCTA	1.000000
35	GCTTCGTGTGGGGATAGTTCTGGCTTGCTAGCABACGGCCCCGATCATGTCCGACGCCACABGAGTGGGBGGACGCGATCACGCA	1.000000
36	GCTTCGTGTGGGGATAGTTCTGGCTTGTGCGCCAACGACCCTGCCCATGCGCGACGACGTGAAAGCGAAAAAACGTGATCACGCA	1.000000
37	GCTTCGTGTGGGAACAGTTCTGGCTTGCTAGCAAGCGGTCGTGCCCATGCGCGACGACGCABAAGCGAAAGGGCGCGATCACGCA	1.000000
38	GCTTCGTGTGGGAACAGTTCTGGCTTGCTAGCABACGGTCCTGCCCATGCGCGACGACGCABAAGCGAAAGGGCGCGATCACGCA	1.000000

APPENDIX E

39	GCTTCGTGTGGGAACAGTTCTGGCTTGCTAGCABACGGTCGTGCCCATGCGCGACGACGCABAAGCGAAAGGGCGCGATCACGCA	1.000000
40	GCTTCGTGTGAGGACAGTTCTGGGCCACTCAGABGCGGTCCTGCCCATGCGCGACGACGTGAGAGTGAAAAGGCACGATCACGCA	1.000000
41	GCTTCGTGTAGGGGCAGTTCTGGCTTGCGCGCCAACGGTCGCGATCATGTCCGACGCCACABGAGTGGGBGGACGCGATCACGCA	1.000000
42	GCTTCGTGCGGTGATAGTTCTGGGTTGCGCGCCAACGGTCCTGCCCATGCGCGACGACGTGAGAGTGGGAGGACGTGACGGCCTA	1.000000
43	GCTTCGGACAAGGGCBTTGTCAGGCCACTCACAAATGGCCCTGCCCTTGTGCGACGACGTGAGAGTGGGAGGACGCGATCACGCA	1.000000
44	GCTTCCTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGCCCCTATCATGCGCGACACCGCGBAAGCGAAAGGGCGTGGCGGCCTA	1.000000
45	GCTTCCTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGCCCCTATCATGCGCGACACCGCGBAAGCGAAAAGGCGTGACGGCCTA	2.000000
46	GCTTCCTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGTCCCGATTATGCGCGACGCCGCGBAAGTGGGAGGACGTGACGGCCTA	1.000000
47	GCTTCCTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGTCCCTATCATGCGCGACACCGCGBAAGCGAAAGGGCGTGGCGGCCTA	2.000000
48	GCTTCCTGTGGGGGCAGTTCTGGCTTGCGCGCCAACGGTCCCTATCATGCGCGACACCGCGBAAGCGAAAAAGTGCAATCACGCA	1.000000
49	GCTTCCTGTGGGGGCAGTTCTGGCTTGCTCACAAACGGCCCTGCCCATGCGCGACGACGCABAAGCGAAAGGGCGCGATCACGCA	1.000000
50	GCTTCCTGTGGGGGCAGTTCTGGCCCACGCACAAATGATCCCGATCATGCGCGACACCGCGBGGGTGGAAGGACGCGATCACGCA	1.000000
51	GCTTCCTGTGGGGGCAGTTCTGGCCCACTCAGABGCGGCCCCGCTCATGCGCGACGACGTGAGAGTGGGAGAACGTGACGGCCTA	1.000000
52	GCTTCCTGTGGGGGCATGTCTGGGTTGCGCGCCAACGGCCCCTATCATGCGCGACACCGCGBAAGCAAAAAGGCGTGACGGCCTA	1.000000
53	GCTTGGTGTGAGGACAGTTCTGAGCCACGCGCCAACGGCCCTGCCCATGCGCGACAATGCGAAAATGGGAAAGCGCAACGGCCTA	1.000000
54	GCTTGGTGCAAGGGTAGTTCTGGCTTGTGCGCCAACGGCCCCGATCATGCGCGACGACGCGBGAGTGGGAAAGCGCAACGGCGCA	1.000000
55	GCAACGTGTGGGGGCATGTCTGGCTTGCTCGCABACGGCCCCGATCATGTCCGACGCCATGBGAGTGGGAGGACGCGATCACGCB	1.000000
56	ACTTCGTGTGGGGGCATGTCTGGGCCACTCAGABGCGGCCCCGCTCATGCGCGACGACGTGAGAGTGGGAGAACGTGACGGCCTA	1.000000
57	ACTTCGTGTGAGGGCAGTTCTGGGTTGCGCGCCAACGGTCCCGATCATGCGCGACACCGCGBAAATGGGAGGACGCGACGGCCTA	1.000000
58	ACTTCGTGTGAGGACAGTTCTGAGCCACTCACAAACGGCCCTGCCCATGCGCGACGACGTGAAAGCGAAAAGACGCGATCACGCA	1.000000
59	ACTTCGTGTGAGGACAGTTCTGAGCCACTCACAAACGGCCCTGCCCATGCGCGAAGACGTGAAAGCGAGAGGACGCGATCACGCA	1.000000
60	ACTTCGTGTGAGGACAGTTCTGAGCCACTCAGAAGCGGTCCTGCCCATGCGCGACGACGTGAGAGTGAAAAGGCACGATCACGCA	1.000000
61	ACTTCGTGTGAGGACAGTTCTGAGCCACTCAGABGCGGCCCTGCCCATGCGCGACGACGTGAGAGTGAAAAGGCACGATCACGCA	1.000000
62	ACTTCGTGTGAGGACAGTTCTGAGCCACTCAGABGCGGCCCTGCCCATGCGCGAAGACGTGAGAGTGAAAAGGCACGATCACGCA	1.000000
63	ACTTCGTGTGAGGACAGTTCTGAGCCACTCAGABGCGGTCCTGCCCATGCGCGACGACGTGAGAGTGAAAAGGCACGATCACGCA	1.000000
64	ACTTCGTGTGAGGACAGTTCTGAGCCACTCAGABGCGGTCGTGCCCATGCGCGACGACGTGAGAGTGAAAAGGCACGATCACGCA	1.000000
65	ACTTCGTGTGAGGACAGTGCTGGCTTGTGCGCCAACGATCGCGATCATATCTAGAGCCGCGBAAGCGAAAGGGCGTGGCGACCCA	1.000000
66	ACTTCGTGCAAGGGCBTGGTCAGCTTGCGCGCCAACGGTCGCTATCATGCGCGACAATGCGBAAATGGGAAAGCGCAGTCACGCA	1.000000
67	ACTTCGTGCAAGGGCBTGGTCAGCTTGCTCAGABGCGGTCGCGCCCATGCGCGACGACGTGAGAGTGAAAAGGCACGACGGCCTA	1.000000
68	ACTTCGGACAAGGGCATTGTCAGGTTGCGCGCCAACGGCCCCGATCAAGCGCAACAATGCGBGAGTGGGAGGACGCGGCGGCCTA	1.000000
69	ACTTCGGACAAGGGCATTGTCAGGCCACTCACAAATGGCCCTGCCCTTGTGCGACGACGTGAGAGTGGGAGGACGCGATCACGCA	1.000000
70	ACTTCGGACAAGGGCATTGTCAGGCCACTCACAAATGGTCCTGCCCTTGTGCGACGACGTGAGAGTGGGAGGACGCGATCACGCA	1.000000
71	ACTTCGGACAAGGGCATGGTCAGCTTGCTCGCABACGGCCCCGATCATATCTAGAGCCGCGBAAGCGAAAGGGCGCGATCGCCTA	1.000000
72	ACTTCGGACAAGGGCBGTTCTGAGCCACTCACAAATGGCCCCGATCATGCGCGACGCCGCGAAAGCGAAAGGGCGCGATCACGCA	1.000000
73	ACTTCGGACAAGGGCBTTGTCAGGTTGCGCGCCAACGGCCCTGATCATGCGCGACGCCACABGAGTGGGAGGACGCGATCACGCA	1.000000
74	ACTTCGGACAAGGGCBTTGTCAGGCCACTCACAAATGGCCCTGCCCTTGTGCGACGACGTGAGAGTGGGBGGACGCGATCACGCB	5.000000
75	ACTTCGGACAAGGGCBTTGTCAGGCCACTCACAAATGGTCCTGCCCTTGTGCGACGACGTGAGAGTGGGAGGACGCGATCACGCA	1.000000
76	ACTTCGGACAAGGGCBTTGTCAGGCCATGCGCCAATGGTCCCTATCTTGTGCGACGACGTGAGAGTGGGBGGACGCGATCACGCB	1.000000
77	ACTTCGGACAAGGGCBTTGTCAGCTTGCGCGCCAACGGCCCCGATCAAGCGCGACAATGTGAGAGTGGGBGGACGCGATCACGCA	1.000000
78	ACTTCGGACAAGGGCBTGGTTGGCTTGCGCGCCAACGGCCCCGATCATGCGCGACGACGCABAAGCGAAAGGGCGCGATCACGCA	1.000000
79	ACTTCGGACAAGGGCBTGGTCAGGCCACTCACAAATGGCCCTGCCCTTGTGCGACGACGTGAGAGTGGGBGGACGCGATCACGCB	2.000000
80	ACTTCGGACAAGGGCBTGGTCAGCTTGCGCGCCAACGGCCCCGATCATGCGCGACACCGCGBGGGTGGGBGGACGCGATCATGCA	1.000000
81	ACTTCGGACAAGGGCBTGGTCAGCCCACTCACAAATGGCCCTGCCCTTGTGCGACGACGTGAGAGTGGGBGGACGCGATCACGCB	1.000000
82	ACATCGTGTAAGGGCATGGTCAGCCCACTCAGAAACGGCCCTGCTCATGCGCGACGCCGCGBGAGTGGGAGGACGTGACGGCCTA	1.000000

APPENDIX F



Appendix G

Minor Allele Frequencies

#	Name	Pos	Obs HET	Pred HET	HWpval	%Geno	Fam Trio	Mend Err	MAF	Alleles
1	SNP1	125	0.407	0.401	1.0	57.4	0	0	0.278	G:A
2	SNP2	181	0.161	0.2	0.631	66.0	0	0	0.113	C:T
3	SNP3	204	0.056	0.054	1.0	76.6	0	0	0.028	T:A
4	SNP4	214	0.028	0.027	1.0	76.6	0	0	0.014	T:A
5	SNP5	236	0.056	0.054	1.0	76.6	0	0	0.028	C:G
6	SNP6	360	0.353	0.327	1.0	72.3	0	0	0.206	G:C
7	SNP7	499	0.289	0.317	0.862	80.9	0	0	0.197	T:G
8	SNP8	504	0.297	0.323	0.896	78.7	0	0	0.203	G:A
9	SNP9	579	0.436	0.393	0.877	83.0	0	0	0.269	T:C
10	SNP10	587	0.474	0.411	0.660	80.9	0	0	0.289	G:A
11	SNP11	599	0.475	0.462	1.0	85.1	0	0	0.362	G:A
12	SNP12	610	0.025	0.025	1.0	85.1	0	0	0.012	G:T
13	SNP13	663	0.075	0.072	1.0	85.1	0	0	0.038	G:A
14	SNP14	690	0.6	0.455	0.103	85.1	0	0	0.35	G:A
15	SNP15	700	0.425	0.362	0.580	85.1	0	0	0.238	C:T
16	SNP16	713	0.4	0.32	0.306	85.1	0	0	0.2	A:T
17	SNP17	830	0.459	0.382	0.476	78.7	0	0	0.257	G:T
18	SNP18	835	0.395	0.317	0.357	80.9	0	0	0.197	T:G
19	SNP19	876	0.415	0.356	0.614	87.2	0	0	0.232	T:G
20	SNP20	953	0.405	0.35	0.647	89.4	0	0	0.226	C:T
21	SNP21	981	0.405	0.35	0.647	89.4	0	0	0.226	T:C
22	SNP22	985	0.405	0.35	0.647	89.4	0	0	0.226	G:A
23	SNP23	1006	0.19	0.172	1.0	89.4	0	0	0.095	G:A
24	SNP24	1173	0.641	0.499	0.164	83.0	0	0	0.474	C:G
25	SNP25	2741	0.564	0.484	0.533	83.0	0	0	0.41	T:C
26	SNP26	2770	0.564	0.484	0.533	83.0	0	0	0.41	T:C
27	SNP27	2777	0.564	0.484	0.533	83.0	0	0	0.41	G:A
28	SNP28	2868	0.41	0.326	0.287	83.0	0	0	0.205	C:T
29	SNP29	2869	0.615	0.5	0.293	83.0	0	0	0.487	G:T
30	SNP30	2907	0.231	0.242	1.0	83.0	0	0	0.141	C:A

APPENDIX G

31	SNP31	2921	0.564	0.484	0.533	83.0	0	0	0.41	G:A
32	SNP32	2932	0.231	0.204	1.0	83.0	0	0	0.115	C:G
33	SNP33	2985	0.59	0.488	0.371	83.0	0	0	0.423	A:C
34	SNP34	3031	0.436	0.341	0.213	83.0	0	0	0.218	A:T
35	SNP35	3043	0.0	0.165	0.003	46.8	0	0	0.091	A:G
36	SNP36	3780	0.311	0.391	0.282	95.7	0	0	0.267	C:T
37	SNP37	3810	0.022	0.022	1.0	95.7	0	0	0.011	G:T
38	SNP38	3813	0.2	0.215	1.0	95.7	0	0	0.122	G:A
39	SNP39	4008	0.644	0.437	0.002	95.7	0	0	0.322	C:T
40	SNP40	4039	0.022	0.022	1.0	95.7	0	0	0.011	C:T
41	SNP41	4042	0.222	0.198	1.0	95.7	0	0	0.111	C:G
42	SNP42	4158	0.512	0.481	0.994	87.2	0	0	0.402	C:T
43	SNP43	4165	0.171	0.232	0.256	87.2	0	0	0.134	G:T
44	SNP44	4960	0.512	0.489	1.0	87.2	0	0	0.427	A:C
45	SNP45	4976	0.537	0.476	0.685	87.2	0	0	0.39	T:C
46	SNP46	4988	0.024	0.024	1.0	87.2	0	0	0.012	C:T
47	SNP47	5153	0.293	0.283	1.0	87.2	0	0	0.171	A:T
48	SNP48	5179	0.049	0.048	1.0	87.2	0	0	0.024	T:A
49	SNP49	5217	0.049	0.048	1.0	87.2	0	0	0.024	G:A
50	SNP50	5279	0.429	0.427	1.0	89.4	0	0	0.31	C:T
51	SNP51	5282	0.238	0.245	1.0	89.4	0	0	0.143	G:C
52	SNP52	5499	0.043	0.043	1.0	97.9	0	0	0.022	C:T
53	SNP53	5514	0.023	0.067	0.071	91.5	0	0	0.035	G:A
54	SNP54	5532	0.044	0.043	1.0	95.7	0	0	0.022	A:G
55	SNP55	5558	0.089	0.085	1.0	95.7	0	0	0.044	C:A
56	SNP56	5653	0.404	0.409	1.0	100.0	0	0	0.287	G:A
57	SNP57	5699	0.66	0.5	0.066	100.0	0	0	0.5	A:A
58	SNP58	5703	0.106	0.101	1.0	100.0	0	0	0.053	C:T
59	SNP59	5810	0.222	0.231	1.0	95.7	0	0	0.133	G:A
60	SNP60	5833	0.617	0.477	0.099	100.0	0	0	0.394	C:T
61	SNP61	6028	0.255	0.282	0.766	100.0	0	0	0.17	G:A
62	SNP62	7054	0.478	0.5	0.940	97.9	0	0	0.5	A:A
63	SNP63	8286	0.349	0.431	0.327	91.5	0	0	0.314	G:A
64	SNP64	8305	0.163	0.187	0.744	91.5	0	0	0.105	A:G
65	SNP65	8324	0.093	0.089	1.0	91.5	0	0	0.047	G:A
66	SNP66	8332	0.326	0.381	0.514	91.5	0	0	0.256	T:C
67	SNP67	8387	0.023	0.023	1.0	91.5	0	0	0.012	G:A
68	SNP68	8417	0.488	0.461	1.0	91.5	0	0	0.36	G:A

APPENDIX G

69	SNP69	8431	0.605	0.493	0.270	91.5	0	0	0.442 G:A
70	SNP70	8489	0.326	0.273	0.566	91.5	0	0	0.163 A:T
71	SNP71	8676	0.216	0.368	0.035	78.7	0	0	0.243 G:A
72	SNP72	8751	0.135	0.171	0.535	78.7	0	0	0.095 G:A
73	SNP73	8791	0.595	0.47	0.235	78.7	0	0	0.378 A:G
74	SNP74	9045	0.0	0.051	0.027	80.9	0	0	0.026 C:T
75	SNP75	9053	0.147	0.185	0.579	72.3	0	0	0.103 G:A
76	SNP76	10086	0.333	0.416	0.343	83.0	0	0	0.295 C:T
77	SNP77	10215	0.051	0.05	1.0	83.0	0	0	0.026 G:A
78	SNP78	10344	0.125	0.117	1.0	85.1	0	0	0.062 A:G
79	SNP79	10391	0.348	0.423	0.343	97.9	0	0	0.304 T:C
80	SNP80	10401	0.356	0.411	0.526	95.7	0	0	0.289 C:G
81	SNP81	10489	0.341	0.442	0.209	93.6	0	0	0.33 A:G
82	SNP82	10533	0.023	0.022	1.0	93.6	0	0	0.011 C:T
83	SNP83	10574	0.386	0.442	0.561	93.6	0	0	0.33 G:C
84	SNP84	10581	0.386	0.442	0.561	93.6	0	0	0.33 C:T
85	SNP85	10622	0.333	0.278	0.546	89.4	0	0	0.167 A:T

Abbreviations Used:

Pos= Position

Obs HET= Observed heterozygosity

Pred HET= Predicted heterozygosity Err

Mend Err= Mendelian error

%Geno= percentage genotype

Fam Tri= Family trio

HWpval= Hardy Weinberg pvalue

MAF= Minor Allele Frequency

Appendix H

Association study using Haploview
Case: Slow lines, Control: Fast lines

#	Name	Assoc Allele	Case,Control Ratio-Counts	Case,Control Frequencies	Chi-square	P value
1	SNP1	G	25:7, 14:8	0.781, 0.636	1.364	0.2428
2	SNP2	T	6:34, 1:21	0.150, 0.045	1.549	0.2133
3	SNP3	T	39:1, 31:1	0.975, 0.969	0.026	0.8726
4	SNP4	A	1:39, 0:32	0.025, 0.000	0.811	0.3677
5	SNP5	C	40:0, 30:2	1.000, 0.938	2.571	0.1088
6	SNP6	C	8:30, 6:24	0.211, 0.200	0.011	0.9151
7	SNP7	G	8:30, 7:31	0.211, 0.184	0.083	0.7732
8	SNP8	G	32:8, 27:7	0.800, 0.794	0.004	0.95
9	SNP9	T	31:11,26:10	0.738, 0.722	0.025	0.8748
10	SNP10	G	30:10,24:12	0.750, 0.667	0.64	0.4238
11	SNP11	G	33:11,18:18	0.750, 0.500	5.355	0.0207
12	SNP12	T	1:43, 0:36	0.023, 0.000	0.829	0.3627
13	SNP13	A	3:41, 0:36	0.068, 0.000	2.55	0.1103
14	SNP14	A	16:28,12:24	0.364, 0.333	0.08	0.7774
15	SNP15	T	13:31,6:30	0.295, 0.167	1.813	0.1781
16	SNP16	A	36:8, 28:8	0.818, 0.778	0.202	0.6531
17	SNP17	G	33:9, 22:10	0.786, 0.688	0.918	0.338
18	SNP18	T	37:7, 24:8	0.841, 0.750	0.967	0.3256
19	SNP19	T	36:8, 27:11	0.818, 0.711	1.327	0.2493
20	SNP20	C	38:8, 27:11	0.826, 0.711	1.588	0.2077
21	SNP21	T	38:8, 27:11	0.826, 0.711	1.588	0.2077
22	SNP22	G	38:8, 27:11	0.826, 0.711	1.588	0.2077
23	SNP23	G	44:2, 32:6	0.957, 0.842	3.162	0.0754
24	SNP24	G	21:21,16:20	0.500, 0.444	0.24	0.6242
25	SNP25	T	27:17,19:15	0.614, 0.559	0.238	0.6255
26	SNP26	T	27:17,19:15	0.614, 0.559	0.238	0.6255
27	SNP27	G	27:17,19:15	0.614, 0.559	0.238	0.6255
28	SNP28	T	10:34,6:28	0.227, 0.176	0.304	0.5816
29	SNP29	T	23:21,15:19	0.523, 0.441	0.511	0.4749
30	SNP30	A	9:35, 2:32	0.205, 0.059	3.362	0.0667

APPENDIX H

31	SNP31	G	27:17,19:15	0.614, 0.559	0.238	0.6255
32	SNP32	C	43:1, 26:8	0.977, 0.765	8.49	0.0036
33	SNP33	A	28:16,17:17	0.636, 0.500	1.461	0.2267
34	SNP34	T	10:34,7:27	0.227, 0.206	0.051	0.8205
35	SNP35	A	22:2, 18:2	0.917, 0.900	0.037	0.8481
36	SNP36	T	18:28,6:38	0.391, 0.136	7.474	0.0063
37	SNP37	T	1:45, 0:44	0.022, 0.000	0.967	0.3254
38	SNP38	A	9:37, 2:42	0.196, 0.045	4.729	0.0297
39	SNP39	T	16:3, 13:31	0.348, 0.295	0.282	0.5951
40	SNP40	C	46:0, 43:1	1.000, 0.977	1.057	0.3039
41	SNP41	C	41:5, 39:5	0.891, 0.886	0.006	0.9406
42	SNP42	T	20:22,13:27	0.476, 0.325	1.947	0.1629
43	SNP43	G	40:2, 31:9	0.952, 0.775	5.55	0.0185
44	SNP44	C	19:25,16:22	0.432, 0.421	0.01	0.9217
45	SNP45	C	19:25,13:25	0.432, 0.342	0.69	0.4063
46	SNP46	C	44:0, 37:1	1.000, 0.974	1.172	0.279
47	SNP47	T	11:33, 3:35	0.250, 0.079	4.214	0.0401
48	SNP48	T	44:0, 36:2	1.000, 0.947	2.374	0.1234
49	SNP49	G	43:1, 37:1	0.977, 0.974	0.011	0.9163
50	SNP50	T	21:25, 5:33	0.457, 0.132	10.281	0.0013
51	SNP51	C	10:36, 2:36	0.217, 0.053	4.613	0.0317
52	SNP52	C	49:1, 41:1	0.980, 0.976	0.016	0.9007
53	SNP53	G	47:1, 36:2	0.979, 0.947	0.637	0.4248
54	SNP54	A	47:1, 41:1	0.979, 0.976	0.009	0.9239
55	SNP55	C	47:1, 39:3	0.979, 0.929	1.35	0.2452
56	SNP56	G	40:1, 27:17	0.800, 0.614	3.97	0.0463
57	SNP57	C	27:23,20:24	0.540, 0.455	0.684	0.4083
58	SNP58	C	50:0, 39:5	1.000, 0.886	6.001	0.0143
59	SNP59	A	10:40, 2:38	0.200, 0.050	4.327	0.0375
60	SNP60	T	20:30,17:27	0.400, 0.386	0.018	0.8926
61	SNP61	A	13:37, 3:41	0.260, 0.068	6.097	0.0135
62	SNP62	A	26:22,20:24	0.542, 0.455	0.697	0.4038
63	SNP63	G	31:11,28:16	0.738, 0.636	1.033	0.3096
64	SNP64	G	6:36, 3:41	0.143, 0.068	1.279	0.2581
65	SNP65	G	42:0, 40:4	1.000, 0.909	4.004	0.0454
66	SNP66	T	32:10,32:12	0.762, 0.727	0.135	0.7129
67	SNP67	A	1:41, 0:44	0.024, 0.000	1.06	0.3032
68	SNP68	G	30:12,25:19	0.714, 0.568	1.99	0.1584

69 SNP69 G 24:18,24:20 0.571, 0.545 0.059 0.8084
70 SNP70 T 7:35, 7:37 0.167, 0.159 0.009 0.9242
71 SNP71 G 35:5, 21:13 0.875, 0.618 6.613 0.0101
72 SNP72 G 39:1, 28:6 0.975, 0.824 4.923 0.0265
73 SNP73 A 28:12,18:16 0.700, 0.529 2.274 0.1316
74 SNP74 C 40:0, 34:2 1.000, 0.944 2.282 0.1309
75 SNP75 G 37:1, 24:6 0.974, 0.800 5.477 0.0193
76 SNP76 C 37:11,18:12 0.771, 0.600 2.591 0.1075
77 SNP77 G 48:0, 28:2 1.000, 0.933 3.284 0.0699
78 SNP78 A 48:0, 27:5 1.000, 0.844 8.0 0.0047
79 SNP79 T 42:8,22:20 0.840, 0.524 10.7790.0010
80 SNP80 C 42:8, 22:18 0.840, 0.550 9.097 0.0026
81 SNP81 A 40:8, 19:21 0.833, 0.475 12.68 4.0E-4
82 SNP82 C 48:0, 39:1 1.000, 0.975 1.214 0.2706
83 SNP83 G 39:9, 20:20 0.812, 0.500 9.643 0.0019
84 SNP84 C 39:9, 20:20 0.812, 0.500 9.643 0.0019
85 SNP85 T 10:36, 4:34 0.217, 0.105 1.884 0.1699
After 10000 permutations
Name Chi Permutation
Square p-value

SNP81 12.68 0.0013
SNP79 10.78 0.0143
SNP50 10.28 0.0401

#10000 permutations performed.
Name Chi Square Permutation p-value
Block 6: TCAGC 11.061 0.0044
Block 6: CGGCT 9.701 0.0097